

Project Documentation

General

- What is the Project purpose?
 - The project's purpose was to glean information from The Association of Tennis Professionals results over the last 10 years. I wanted to see what players were playing most often from names to the rankings. I also wanted to let newcomers know what tournaments were the biggest. I also wanted to be able to see who was playing who, so I made a network graph where you can see opponents from different tournaments.
- Data description + How was the data collected?
 - I have two datasets, one from 2013-2023 and one from 2000-2014. The majority of the project focused on the more recent results and more in-depth information on the data sets can be found under the "Data Sets and Descriptions" folder. The data was collected by the ATP tour and aggregated into the two data sets. Both data sets have probability data from different betting companies, but that data was not used in the visualizations.
- Who are the users that this Shiny App was made for?
 - This project was designed for people who play and are interested in tennis. As a tennis player myself, I created visualizations that I found interesting in the hopes that other players like me would find the information useful.
- Questions I am trying answer
 - I wanted to know what players played the most matches on the ATP tour.
 - Word Cloud
 - I wanted to show newcomers what the biggest tournaments were.
 - Word Cloud
 - I wanted to see what ranking you would need to compete in ATP tournaments as opposed to the entry tournaments like challengers.
 - Histogram
 - I wanted to be able to compare the rankings of players over the time period of the data.
 - Player Ranking over Time
 - I wanted to know who has played who at different tournaments.
 - Who Played Who (network graph)
- Insights
 - From the Word Cloud I learned there are certain players that I realized were more prolific than I knew about such as Grigor D. I also learned that Djokovic N. was much more dominant than I realized both from 2000-2014 but also 2013-2023.
 - From the histogram I learned there is a significant drop off after the top 20, but also much fewer appearances by anyone with a rank lower than 300.
 - By comparing the rankings in the "Player ranking over time" chart you can see things like drop in ranking due to injury and how the top players like Djokovic, Nadal, Federer, and Alcaraz compare to each other over time and when each grabbed the top spot and how long they kept it.

- Lastly from the network I found that Federer and Djokovic have never played at the China Open before.
- What needs improvement? (Wish list)
 - I want to further integrate the results from the older data set (2000-2014) into more of the graphs.
 - I want to use the location data from the older data set to plot the locations of tournaments on a map.
 - Creating a good-looking map was proving to be difficult.
 - I had wanted to include match data on the network when you hovered over the nodes.
 - The igraph structure used to make the network was difficult to work with though and it was not possible to plot a graph with multiple matches with the same person in ggnet2.
 - I also wanted to look into player upsets and see which players had the most or least upsets.
- Sources or References (with appropriate credits)
 - Specific uses listed in code with associated numbers.

[1]

codersgram9, “Plot lines from a list of dataframes using ggplot2 in R,” GeeksforGeeks. Accessed: Oct. 10, 2023. [Online]. Available: <https://www.geeksforgeeks.org/plot-lines-from-a-list-of-dataframes-using-ggplot2-in-r/>

[2]

dissfya, “ATP Tennis 2013 - 2023,” Kaggle. Accessed: Oct. 10, 2023. [Online]. Available: <https://www.kaggle.com/datasets/dissfya/atp-tennis-2013-2023>

[3]

J. Goblet, “ATP Men’s Tour,” Kaggle. Accessed: Oct. 10, 2023. [Online]. Available: <https://www.kaggle.com/datasets/jordangoblet/atp-tour-20002016>

[4]

“briatte/ggnet: vignettes/ggnet2.Rmd.” Accessed: Oct. 10, 2023. [Online]. Available: <https://rdrr.io/github/briatte/ggnet/f/vignettes/ggnet2.Rmd>

[5]

“Plotly: Low-Code Data App Development.” Accessed: Oct. 10, 2023. [Online]. Available: <https://plotly.com/>

[6]

“Shiny for R Gallery.” Accessed: Oct. 10, 2023. [Online]. Available: <https://shiny.posit.co/r/gallery/>

[7]

“table function - RDocumentation.” Accessed: Oct. 10, 2023. [Online]. Available: <https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/table>

[8]

“ggnet2: network visualization with ggplot2.” Accessed: Oct. 10, 2023. [Online]. Available: <https://briatte.github.io/ggnet/>

[9]

“igraph – Network analysis software.” Accessed: Oct. 10, 2023. [Online]. Available: <https://igraph.org/>

Process

- I started out by creating a couple visualizations following the shiny gallery examples:
 - Word Cloud
 - Needed to make a frequency table of each element to display
 - Added tournament names
 - Added player names
 - Included a way to resize the graph
 - Later I would add in another older data set to see what players fell off the list of top players
 - Histogram
 - Added in the player rankings to see what ranks were most common in results
 - I changed the bin size options
 - Changed up the color to match a tennis court green and blue
- Next I decided to create a line plot
 - At first I included one player's results
 - I had to filter by the player's name
 - Later I added functionality to use multiple names
 - Had to use player names as a color indicator
 - ggplot has some useful parameters for multi-plotting
 - used selectizeInput() from shiny to get multiple inputs
- My final figure was a network
 - I first created a network of all opponents of a player
 - Too messy
 - Could not get hovering to work
 - I went to Vivero and they suggested I use a second filtering
 - I decided to use tournaments to filter and make smaller networks and it worked well
 - I added color
 - Green and blue to match a tennis court
- I dabbled a little in creating a map
 - I tried a bit to make a map of tournament location using the older 2000-2014 data set locations list
 - The maps look poor and I ultimately scrapped the idea due to time constraints

Design Decisions:

Word Cloud

- What:
 - The names of players and tournaments in a table that was manipulated to count the frequency of each occurrence in the data set. An older data set with player names was also included to see comparisons between time frames.
- Why:
 - To understand which players played more ATP matches and which tournaments had bigger draws and thus more matches.
- How:
 - By placing each name by size in accordance with the frequency in which each one appears. Different colors were added to allow for easier identification of size groupings and sliders were used to allow for better selection of data and the ability to resize the graph.

Histogram

- What:
 - The ranks of players in each match contained in the dataset as a table.
- Why:
 - To understand the relationship between ranking and how often a player appears on the tour. Do some top players forgo tournaments more often? How quickly does the appearance of certain rankings drop off?
- How:
 - By placing each ranking result in a bin and counting the occurrences. There is also an optional “run” figure that displays a line for each rank along the x-axis. Finally, a density curve is added to show how the data is behaving. Color was added to make the graph more appealing.

Player Ranking over Time

- What:
 - The rankings of different players over the course of 2013-2023.
- Why:
 - To understand how players have moved up and down the rankings over time and how they stack up against other players.
- How:
 - Using a line graph, the players rankings over the period are displayed all at once and more detailed information can be seen by hovering over the line. Different colors are used to differentiate between different players, but you can also tell which line belongs to each player by hovering over a line and more information will be displayed. Zooming, panning, and comparing rankings on different lines is also possible.

Who Played Who

- What:
 - The names of players and their opponents along with what tournament the match happened at. These results are filtered to just get the opponents historically played against as a particular tournament.
- Why:
 - To understand who a player has and has not played at a particular tournament.
- How:
 - A simple graph with connections between each opponent. You can zoom and pan around for easier viewing of denser graphs. Colors were added to make the graph more appealing.

Time Log:

starting Github repository

-15min (8/29/2023)

-15min (9/7/2023)

Selecting Dataset

-20min (8/29/2023)

-10min (9/3/2023)

Describing Dataset

-30min (9/17/2023)

-30min (9/24/2023)

Starting R file

-10min (9/10/2023)

Making Shiny Apps

-5hrs (9/24/2023)

histogram/ word cloud creation

-3.5hrs (10/1/2023)

line graph implementation

-7.75hrs (10/8/2023)

Network initial implementation

-5.5hrs (10/9/2023)

Network improvement, map testing with extra data set, and documentation

-7.5hrs (10/10/2023)

network, extra data to word cloud, documentation

Time Chart

(see the R-notebook for interactive view)

